Convergence in distribution and probability

Convergence in distribution

Convergence in distribution, denoted by

$$X_n \xrightarrow{d} X,$$

means that $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$. (In fact, this is only required to happen for values of x for which $F(x)$ is continuous.) The *distribution* of X_n approaches that of X, but the value of X_n need not be close to the value of X.

Direct application of the definition is sometimes hard. Often much easier to show the equivalent condition $\varphi_{X_n}(u) \to \varphi_X(u)$ with $\varphi_X(u)$ continuous at $u = 0$. Using this technique we have shown that $b(p,n) \xrightarrow{d} N(n, npq)$, which simplifies the computation of binomial probabilities for large n.

Convergence in probability

Motivation

Sometimes what we want to formalize is not just that X_n has a *distribution* which approaches that of X, but rather that *the value* of X_n approaches that of X as $n \to \infty$.

> EXAMPLE. If we have a regular coin ($P(\text{heads}) = P(\text{tails}) = 0.5$) we have the feeling that in a very long series of throws,
>
> $$X_n = \frac{\text{Total heads in } n \text{ throws}}{n} \tag{1}$$
>
> would approach $p = 0.5$.

We might be tempted to resort to the ordinary notion of limit we learned in high school Calculus, and write that

$$\lim_{n\to\infty} X_n = p. \tag{2}$$

In fact, equations such as (2) are sometimes given as a "definition" of the probability of the event "heads"! We will see that this is misguided.

Why the ordinary notion of limit fails with random variables

Let's recall what limit means in Calculus. If we have a sequence $\{a_n\}_{n=1}^{\infty}$ whose n-th term is given by,

$$a_n = \frac{1}{n},$$

we say that $\lim_{n\to\infty} a_n = 0$ to mean that a_n will be as close to zero as we wish, if we take a large enough n. A formal way to express that is:

Ordinary limit

> $\lim_{n\to\infty} a_n = \ell$ if for any $\epsilon > 0$ there exists $N(\epsilon)$ such that whenever $n > N(\epsilon)$ we have $|a_n - \ell| < \epsilon$.

Now we may ask ourselves: can we say that the sequence of random variables (1) converges to p in this sense?

A little thought will reveal us that the answer is "No." No matter how large is n, it *could* be that $X_n = 1$ (because all throws of the coin have shown heads, and extremely unlikely event, but nonetheless possible).

So, no, we cannot guarantee that X_n will be close to p for $n > N(\epsilon)$, however large we choose $N(\epsilon)$.

From the limit of Calculus to the limit in probability

We can extend to random sequences the previous notion of limit in various ways. Probably the most natural is, expressed in plain words:

> X is the limit *in probability* of X_n if for large enough n we can make $|X_n - X|$ arbitrarily small *with arbitrary high probability.*

The only difference with the ordinary limit of Calculus is that we no longer require that $|X_n - X|$ will be small with certainty, which we cannot do, but rather with probability that can be made as close to 1 as we desire.

The notation for convergence in probability of X_n to X is $X_n \xrightarrow{p} X$, or alternatively, $\text{plim} X_n = X$. The formal definition would read as follows:

Limit in probability

> $X_n \xrightarrow{p} X$ if for any $\epsilon > 0, \eta > 0$ there is $N(\epsilon, \eta)$ such that whenever $n > N(\epsilon, \eta)$ then $P(|X_n - X| < \epsilon) \geq 1 - \eta$.

An equivalent way of expressing the same thing is:

$$\text{plim} X_n = X \quad \text{if for any } \epsilon > 0 \ \lim_{n \to \infty} P(|X_n - X| < \epsilon) = 1.$$

How to verify convergence in probability

We will use basically two methods.

Direct use of the definition

Rarely, we will be able to check convergence in probability directly from the theory. Problem 1 is an example. We have to check that X_n will be with high probability close to the assumed limit for large enough n.

Using Tchebycheff inequality

To establish convergence in probability of X_n to m we must show that for any $\epsilon > 0$, $\eta > 0$ we can find n so that:

$$P(|X_n - m| < \epsilon) \geq 1 - \eta \tag{3}$$

Remember that for any random variable Z with first two moments, it is true that:

$$P(|Z - m| < k\sigma) \geq 1 - \frac{1}{k^2}. \tag{4}$$

If X_n is a sequence of random variables with common mean m and respective variances approaching zero, $\sigma_n^2 \to 0$, using (4) we have:

$$P(|X_n - m| < k\sigma_n) \geq 1 - \frac{1}{k^2}. \tag{5}$$

Using (5) is a simple matter to show (3) in an easy two-step procedure.

1. Whichever $\eta > 0$ may be, we can take k_* large enough so that $1/k_*^2 < \eta$. For that value of k_*, equation (5) yields:

$$P(|X_n - m| < k_*\sigma_n) \geq 1 - 1/k_*^2 \geq 1 - \eta \tag{6}$$

2. Since $\sigma_n^2 \to 0$, whichever $\epsilon > 0$ may be and for any k_* set in the previous step, w can find n large enough so that $k_*\sigma_n < \epsilon$. For such n, (6) yields:

$$P(|X_n - m| < k_*\sigma_n < \epsilon) \geq 1 - 1/k_*^2 \geq 1 - \eta, \tag{7}$$

   hence (3). It follows that $X_n \xrightarrow{p} m$.

We rarely go through the reasoning above; rather, we have proven this general result that we can apply directly:

> **Theorem.** Any sequence of random variables X_n with common mean m and variances $\sigma_n^2 \to 0$ converges in probability to m.

Law of large numbers

The theorem above covers a very common situation. For instance, we might have a population with unknown mean m and variance σ^2 and take a sample $Z_1, Z_2, \ldots, Z_n$ of size n. If we compute the average

$$\overline{Z}_n = \frac{Z_1 + Z_2 + \ldots + Z_n}{n}$$

it is clear that for $E[\overline{Z}_n] = m$ and $\sigma^2_{\overline{Z}_n} = \sigma^2/n$, so $\sigma^2_{\overline{Z}_n} \to 0$ as n increases. Hence, $\sigma^2_{\overline{Z}_n}$ fulfills the conditions of the theorem and we can claim that $\overline{Z}_n \xrightarrow{p} m$. This is a so-called *law of large numbers*, of which there are many varieties. Its significance is that if we are interested in approximating the value of m, we are confident that taking a large enough sample will take us close to m with high probability.

Law of large numbers

There can be convergence in probability even if the conditions in the theorem do not hold, or convergence can be to something else than the mean. Some problems below illustrate these points.

Convergence in probability is the useful concept, underlying such important notions as consistency (later in the course and in Econometrics next year). Convergence with probability one (or "almost sure convergence") and other so-called strong convergences have little additional practical implications for our applications.

Problems.

1. If we take values from a $U(a, b)$ and compute the maximum of them,

$$X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$$

show directly that $X_{(n)} \xrightarrow{p} b$. (Hint: Do not compute variances or try to use Tchebycheff, use just the definition of convergence in probability. What would be the probability that *all* values $X_1, X_2, \ldots, X_n$ are below $b - \epsilon$ if n grows very large?).

Direct proof of convergence in probability

2. We throw repeatedly a coin with probability p of giving heads. In the i-th throw $X_i = 1$ if we get heads and $X_i = 0$ if not. We keep computing the averages of the first two, three, $\ldots, n$ throws:

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n};$$

What does $\overline{X}_n$ converge to in distribution? In probability?

Binomial frequency converges to true p

3. Consider the sequence of random variables defined as:

$$X_n = \begin{cases} n & \text{with probability } \frac{1}{n} \\ 0 & \text{with probability } 1 - \frac{1}{n} \end{cases}$$

Does it converge in probability? To what? What is the mean of each of the X_n?

Convergence in probability but not to the mean

4. Consider $Z_n \sim \mathcal{P}(\lambda = 1/n)$. Define $X_n = nZ_n$. Show that:

   (a) $E[X_n] = 1$ for all n.

   (b) X_n converges in probability to 0 (although the common mean is 1!).

Convergence in probability but not to the common mean

5. The random variables X_n have the following distribution:

$$X_n \sim N(0, \sigma^2 = 1/n)$$

Does X_n converge in distribution to some variable X? In probability? To what? Is there some value x for which $F_{X_n}(x) \not\to F_X(x)$?

Convergence in distribution

6. If X_n has distribution function

$$F_{X_n}(x) = 1 - \left(1 - \frac{x}{n}\right)^n \qquad 0 < x \leq n$$

then $X_n \xrightarrow{d} X$ with $F_X(x) = 1 - e^{-x}$ (a particular case of the so-called exponential distribution; the general exponential has distribution function $F_X(x) = 1 - e^{-\lambda x}$).

Convergence in distribution

7. Consider $X_n \sim \exp(n)$, i.e. $F_{X_n}(x) = 1 - e^{-nx}$. What does X_n converge to in distribution? In probability?

$X_n \xrightarrow{d} X$ and $X_n \xrightarrow{p} X$.

8. Consider $X_n = X + Y_n$ where $E[Y_n] = 1/n$ and $\text{Var}(Y_n) = \sigma^2/n$. Show that $X_n \xrightarrow{p} X$. (Hint: Take n_1 such that $1/n_1$ is within $\epsilon/2$ of zero. Now, take n_2 such that Y_{n_2} is within $\epsilon/2$ of its mean with probability $1 - \eta$. What happens for $n \geq \max(n_1, n_2)$?)

9. Let $X_i \sim U(0,1)$. Study the convergence in probability of:
   (a) $\{X_i\}$ as $i \to \infty$.
   (b) $\{X_i/i\}$ as $i \to \infty$.

10. In general, $X_n \xrightarrow{p} X$ implies $X_n \xrightarrow{d} X$, but not the other way around. However, when the limit is a constant c, then $X_n \xrightarrow{d} c$ implies $X_n \xrightarrow{p} c$. Give a short proof or an argument to that effect.

Sometimes $X_n \xrightarrow{d} X \Rightarrow X_n \xrightarrow{p} X$.

11. Explain why the statement "Probabilities are the numbers to which relative frequencies would converge in a long series of experiments" cannot be seriously taken as a definition of probability.

Reading. For problem 9 you may want to watch the video https://www.youtube.com/watch?v=eb-eRduYwZY.

You have many other resources in Internet, including https://www.probabilitycourse.com/chapter7/7_2_5_convergence_in_probability.php whence some of the problems above have been taken. Just type "convergence in probability" and you will be swamped with more references than you can digest in a life time.

Regarding books, you can find basic definitions in [5], Chapter 25. Many other books cover these topics, concisely as [1] or in much more detail (and level) than needed in this course, like [2]. In [3] you may also find some examples. Much more comprehensive, and well over the level of this course is [4], a collection of examples and counterexamples which helps to understand the meaning and logic of definitions, and the subtle way in which different convergences differ from one another.

References

[1] Robert B. Ash. *Basic Probability Theory (Dover Books on Mathematics)*. Dover Publications, 2008.

[2] P. Billingsley. *Probability and Measure*. John Wiley and Sons, New York, second edition, 1986.

[3] A. Garín and F. Tusell. *Problemas de Probabilidad e Inferencia Estadística*. Ed. Tébar-Flores, Madrid, 1991.

[4] J. P. Romano and A. F. Siegel. *Counterexamples in Probability and Statistics*. Wadsworth and Brooks/Cole, Monterrey, California, 1986.

[5] A. Fz. Trocóniz. *Probabilidades. Estadística. Muestreo.* Tebar-Flores, Madrid, 1987.

Soluciones abreviadas

1. La probabilidad de que $X_1, \ldots, X_n$ estén todos a la izquierda de $(b-\epsilon)$ es

$$\left[\frac{b-\epsilon-a}{b-a}\right]^n$$

que converge a cero cuando $n \to \infty$. Por lo tanto, la probabilidad del suceso complementario (que *algún* X_i, y por tanto $X_{(n)}$, estén a la derecha de $b-\epsilon$), tiende a 1.

2. Para cualquier n, $E[\overline{X}_n] = p$. Además $\text{Var}(\overline{X}_n) = pq/n \to 0$. Por tanto, problema standard en que la convergencia en probabilidad a p se sigue de la aplicación de Tchebycheff. En distribución, converge a una normal $N(p, pq/n)$ (teorema central del límite).

3. El ejemplo más simple de una sucesión de v.a. que converge en probabilidad, pero no a la media común a todas ellas. $E[X_n] = 1$, pero $X_n \xrightarrow{p} 0$.

4. Cuando $\lambda \to 0$, $Z_n \xrightarrow{p} 0$ (la probabilidad del valor 0 en una Poisson de parámetro λ es $e^{-\lambda}$). Por tanto, también $X_n \xrightarrow{p} 0$. Sin embargo, $E[X_n] = nE[Z_n] = n(1/n) = 1$. No es aplicable la acotación de Tchebycheff para probar convergencia a la media porque las varianzas σ_n^2, lejos de tender a cero, crecen hacia infinito: la distribución de X_n se concentra en cero, pero con valores de probabilidad pequeña "muy lejos", que hacen crecer la varianza al crecer n.

5. Sí converge en distribución, aunque no a una distribución continua. Las funciones de distribución de X_1, X_2, etc. tienen la forma ilustrada en la Figura 1.

   Las funciones de distribución de $X_1, \ldots, X_n$ convergen a la función de distribución (en rojo) de una variable causal, tomando el valor 0 . X_n converge también en probabilidad (a cero). La convergencia $F_{X_n}(x) \to F_X(x)$ se verifica (como debe) en todo punto de continuidad de $F_X(x)$, pero no en $x = 0$ (todas las $F_{X_n}(x)$ toman valor 0.5 ahí); no importa.

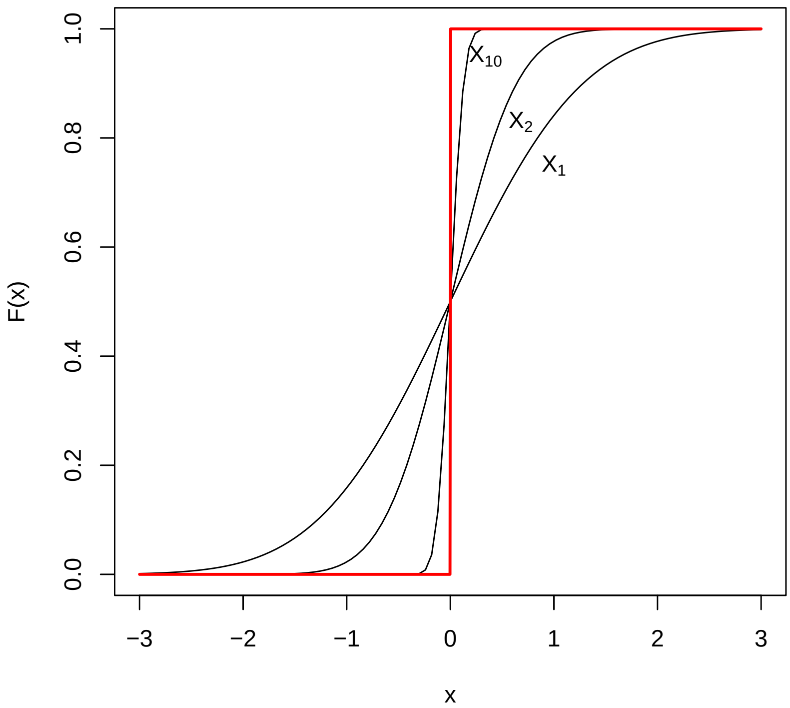
6. Basta tomar límite y recordar que $\lim_{n\to\infty}(1 + a/n)^n = e^a$.

7. La función de distribución es $F_{X_n}(x) = 1 - e^{-nx}$. Para cualquier $x > 0$, esta expresión converge a 1 cuando $n \to \infty$. Por tanto, la convergencia es a una distribución causal que otorga toda la probabilidad al punto $x = 0$. La convergencia en probabilidad es también a dicho valor.

8. Utilizando la ayuda, se trata de tomar un n que verifique simultaneamente estas dos cosas: a) Que el valor medio de Y_n este a $\epsilon/2$ de cero, y b) Que Y_n este a $\epsilon/2$ de su valor medio. Cuando estas dos condiciones se verifican, Y_n está a ϵ de cero, y X_n a ϵ de X.

9. (a) No converge. (b) Converge a cero en probabilidad.

Figure 1: Convergencia en distribución a una causal



10. Puede razonarse sobre la Figura 1. Si la convergencia en distribución es a una causal que concentra toda la probabilidad en el punto c, entonces, considerando un entorno $(c-\epsilon, c+\epsilon)$, términos X_n de la sucesión para n suficientemente grande tendrán distribuciones que "entrarán" en $(c-\epsilon, c+\epsilon)$ con valor menor a $\eta/2$ y "saldran" con valor superior a $1-\eta/2$. Por consiguiente, la probabilidad del intervalo $(c-\epsilon, c+\epsilon)$ será mayor que $1-\eta$ para tal n.

11. El modo en que las frecuencias relativas convergen a las probabilidades es… en probabilidad, lo que haría la definición circular. (No puede hacerse uso del concepto definido en su propia definición.)

Pautas docentes

1. Conviene que entiendan bien el problema 1, pues ejemplos basados en la distribución uniforme (y en concreto el de $\overline{X}_n$ como estimador máximo-verosímil de b) aparecen más adelante.

2. Hacedles notar en el problema 3 que este fenómeno de convergencia a algo diferente de la media común es frecuente cuando con una probabilidad tendente a cero la variable X_n "escapa" hacia ∞. Una variación aún más espectacular de este problema se presenta cuando

$$X_n = \begin{cases} n^2 & \text{with probability } \frac{1}{n} \\ 0 & \text{with probability } 1-\frac{1}{n} \end{cases}$$

¡X_n converge a cero en probabilidad, pero su media diverge hacia ∞!

3. No dejéis de explayaros al interpretar el problema 2. "¡O sea que, aunque no conozcamos p, podemos aproximarnos todo lo que queramos con probabilidad arbitrariamente cercana a 1 simplemente tomando más muestra! Claramente, queremos que cualquier método que diseñemos para aproximar parámetros desconocidos tenga esta propiedad." (Volveremos sobre el particular al hablar de consistencia.)

4. En el pasado hemos mencionado a veces convergencias fuertes (convergencia casi segura y convergencia en media cuadrática). En la situación nuestra, creo que es mejor concentrarse en las dos convergencias débiles, que son las que luego encuentran. Olvidemos las fuertes.

5. Faltaría una ilustración del uso de $\varphi_{X_n}(u) \to \varphi_X(u)$ para establecer $X_n \xrightarrow{d} X$. Una tal ilustración aparece como teoría en uno de los handouts previos (handout 3, binomial a normal). No tengo ningún buen ejemplo alternativo que no sea trivial.